

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2002年 9月17日

出 願 番 号

Application Number:

特願2002-269941

[ST.10/C]:

[JP2002-269941]

出 願 人

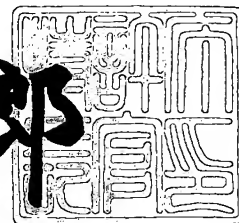
Applicant(s):

株式会社デンソー

2003年 6月20日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特2003-3048549

【書類名】 特許願

【整理番号】 PNID4141

【提出日】 平成14年 9月17日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 13/00

【発明者】

 【住所又は居所】 愛知県刈谷市昭和町 1 丁目 1 番地 株式会社デンソー内

 【氏名】 鈴木 竜一

【発明者】

 【住所又は居所】 愛知県刈谷市昭和町 1 丁目 1 番地 株式会社デンソー内

 【氏名】 笹木 美樹男

【特許出願人】

 【識別番号】 000004260

 【氏名又は名称】 株式会社デンソー

【代理人】

 【識別番号】 100082500

 【弁理士】

 【氏名又は名称】 足立 勉

 【電話番号】 052-231-7835

【手数料の表示】

 【予納台帳番号】 007102

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

 【包括委任状番号】 9004766

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声対話装置及びプログラム

【特許請求の範囲】

【請求項 1】 ユーザから対話のための音声入力となされると、該入力内容を音声認識する認識手段と、

ユーザとの対話内容に応じた複数のシナリオと、該各シナリオに沿った発話対象語を予め記憶した記憶手段と、

前記認識手段による認識に応じて、前記記憶手段に記憶された発話対象語の中からユーザに向けた発話語を選択する選択手段と、

該選択手段によって選択された発話語を音声により出力する出力手段と、

を備え、ユーザとの間で対話を行う音声対話装置であって、さらに、

前記記憶手段に記憶された発話対象語の中に、ユーザとの対話内容に応じた発話対象語がない場合に、前記選択手段にユーザに該対話内容の答えを問い返すための発話語を選択させ、該問い返しに対してユーザから入力された対話内容の答えに基づき、該対話内容に応じたシナリオを学習し、該シナリオと該各シナリオに沿った発話対象語を前記記憶手段に新たに記憶させる学習手段を備えたことを特徴とする音声対話装置。

【請求項 2】 請求項 1 に記載の音声対話装置において、さらに、

前記学習手段により学習された対話情報に基づき、前記記憶手段において、該対話内容についての音声対話に必要なシナリオ、対話辞書、認識辞書を、自動的に更新する更新手段を備えたことを特徴とする音声対話装置。

【請求項 3】 請求項 2 に記載の音声対話装置において、

前記学習手段により学習された対話情報について、複数回交わされた同種の対話内容の履歴を記憶する第 2 の記憶手段を備え、

該第 2 の記憶手段が記憶した対話情報に基づき、同種の対話内容について互いに不整合なシナリオがある場合には、前記更新手段は、対話確率の高いシナリオに順次変更して更新することを特徴とする音声対話装置。

【請求項 4】 請求項 3 に記載の音声対話装置において、

前記更新手段は、前記不整合なシナリオ間で、その対話内容についての前記対

話確率が等しい場合は、先に出現したものを優先することを特徴とする音声対話装置。

【請求項 5】 ユーザから対話のための音声入力となされると、該入力内容を音声認識する認識手段と、

ユーザとの対話内容に応じた複数のシナリオと、該各シナリオに沿った発話対象語を予め記憶した記憶手段と、

前記認識手段による認識に応じて、前記記憶手段に記憶された発話対象語の中からユーザに向けた発話語を選択する選択手段と、

該選択手段によって選択された発話語を音声により出力する出力手段と、

を備え、ユーザとの間で対話を行う音声対話装置であって、

前記記憶手段は、同じ意味内容の対話について複数のバリエーションのシナリオ及び対話対象語を記憶し、

前記選択手段は、ユーザとの音声対話における該ユーザの応答内容に応じて、前記記憶手段から選択する発話語を変化させることを特徴とする音声対話装置。

【請求項 6】 請求項 5 記載の音声対話装置において、さらに、ユーザからの音声入力に基づき、該ユーザの感情を識別する識別手段を備え、

前記選択手段は、該識別手段により識別されたユーザの感情情報に応じて、前記出力手段が出力する発話語の語調を変化させるように、前記記憶手段から発話語を選択することを特徴とする音声対話装置。

【請求項 7】 請求項 6 記載の音声対話装置において、

前記識別手段は、前記ユーザの感情情報が当該音声対話装置に対してのものか、又は一般的なことに対してのものかを識別し、

前記選択手段は、該識別手段による識別結果に応じて、前記出力手段が出力する発話語の語調を変化させるように、前記記憶手段から発話語を選択することを特徴とする音声対話装置。

【請求項 8】 請求項 5 記載の音声対話装置において、さらに、ユーザからの音声入力に基づき、方言を識別する識別手段を備え、

前記選択手段は、該識別手段により識別された方言に応じて、前記出力手段が出力する発話語の語調を変化させるように、前記記憶手段から発話語を選択する

ことを特徴とする音声対話装置。

【請求項 9】 請求項 8 記載の音声対話装置において、

前記選択手段は、該識別手段により識別された方言に応じて、対話の話題が該方言にかかる土地柄にちなんだものになるように、前記記憶手段から発話語を選択可能に構成されたことを特徴とする音声対話装置。

【請求項 10】 請求項 5 記載の音声対話装置において、さらに、ユーザからの音声入力に基づき、その言語を識別する識別手段を備え、

前記選択手段は、該識別手段により識別された言語に応じて、前記出力手段が出力する発話語の語調を変化させるように、前記記憶手段から発話語を選択することを特徴とする音声対話装置。

【請求項 11】 請求項 10 記載の音声対話装置において、

前記選択手段は、該識別手段により識別された言語に応じて、対話の話題が該言語にかかる国にちなんだものになるように、前記記憶手段から発話語を選択可能に構成されたことを特徴とする音声対話装置。

【請求項 12】 請求項 1～11 のいずれかに記載の音声対話装置において、さらに、ユーザからの音声入力に基づき、その声質から該ユーザの属性を判定する判定手段を備え、

前記選択手段は、該判定手段により判定された属性に応じて、前記出力手段が出力する発話語の声質を変化させるように、前記記憶手段から発話語を選択することを特徴とする音声対話装置。

【請求項 13】 請求項 1～12 のいずれかに記載の音声対話装置において、さらに、ユーザの姿態を撮像して画像認識し、該ユーザの属性を判定する判定手段を備え、

前記選択手段は、該判定手段により判定された属性に応じて、前記出力手段が出力する発話語の声質を変化させるように、前記記憶手段から発話語を選択することを特徴とする音声対話装置。

【請求項 14】 請求項 12 又は請求項 13 に記載の音声対話装置において、

前記選択手段は、該判定手段により判定された属性に応じて、対話の話題が該属性にちなんだものになるように、前記記憶手段から発話語を選択可能に構成さ

れたことを特徴とする音声対話装置。

【請求項 1 5】 ユーザから対話のための音声入力となされると、該入力内容を音声認識する認識手段と、

ユーザとの対話内容に応じた複数のシナリオと、該各シナリオに沿った発話対象語を予め記憶した記憶手段と、

前記認識手段による認識に応じて、前記記憶手段に記憶された発話対象語の中からユーザに向けた発話語を選択する選択手段と、

該選択手段によって選択された発話語を音声により出力する出力手段と、

を備え、ユーザとの間で対話を行う音声対話装置であって、さらに、

ユーザの顔画像を撮像し、その唇の動きに基づくリップリーディングにより画像認識する画像認識手段を備え、

前記認識手段は、前記画像認識手段による画像認識を併用して前記音声認識を行うことを特徴とする音声対話装置。

【請求項 1 6】 請求項 1 ～ 1 5 のいずれかに記載の音声対話装置において、ユーザと対話するロボットとして構成されたことを特徴とする音声対話装置。

【請求項 1 7】 請求項 1 5 に記載の音声対話装置において、前記ユーザの顔画像を撮像する目を備えたロボットとして構成され、

前記画像認識手段は、該目により撮像された顔画像から、前記ユーザがロボットの正面を向いているか否かを判定し、

前記認識手段は、前記画像認識手段により前記ユーザがロボットの正面を向いていると判定された場合にのみ、前記音声認識を行うことを特徴とする音声対話装置。

【請求項 1 8】 請求項 1 7 記載の音声対話装置において、

前記ロボットの目が四方を見渡せるように、その頭部周囲に複数設けられ、

前記画像認識手段は、該複数の目のいずれかにより撮像されたユーザの顔画像により、前記ユーザがロボットの方向を向いているか否かを判定し、

前記認識手段は、前記画像認識手段により前記ユーザがロボットの方向を向いていると判定された場合にのみ、前記音声認識を行うことを特徴とする音声対話装置。

【請求項 1 9】 請求項 1 7 又は請求項 1 8 に記載の音声対話装置において、前記認定手段を構成するロボットの耳がその頭部周囲に複数設けられ、前記認識手段は、該複数の耳に入力される音声レベルに基づき、前記ユーザがロボットの方向を向いているか否かを判定し、前記ユーザがロボットの方向を向いていると判定された場合にのみ、前記音声認識を行うことを特徴とする音声対話装置。

【請求項 2 0】 請求項 1 8 又は請求項 1 9 に記載の音声対話装置において、さらに、前記画像認識手段又は前記認識手段により前記ユーザと前記ロボットが向き合っていないと判定された場合に、該ロボットが前記ユーザに向き合うように該ロボットを動作させるように構成されたことを特徴とする音声対話装置。

【請求項 2 1】 請求項 1 ～ 2 0 のいずれかに記載の音声対話装置の前記各手段としてコンピュータを機能させるためのプログラム。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、ユーザとの間で音声対話を行うための音声対話装置に関する。

【0 0 0 2】

【従来の技術】

従来より、例えばカーナビゲーションシステムにおいてレストラン等の目的地の位置情報を音声により問い合わせたりする情報検索のための装置、音声対話を通じてユーザを楽しませる娯楽用の装置等の音声対話装置が知られている。特に近年では、こうした音声対話においてユーザとの間で自然な対話を実現するために、対話のためのシナリオを予め複数用意してユーザの発話に対応する音声対話装置が提案されている（例えば、特許文献 1 参照）。

【0 0 0 3】

【特許文献 1】

特開 2 0 0 1 - 3 5 7 0 5 3 号公報

【0 0 0 4】

【発明が解決しようとする課題】

しかしながら、上記特許文献 1 の音声対話装置では、ユーザの発話に対してシステム側の応答が予め対応づけられており、ユーザの質問等の発話に対して予め決められたことしか答えることができなかった。このため、答えを知らない場合には応答することができず、対話を中断するか話題を変えるなどの手段をとるしかなく、知的な対話を行うという観点からは十分ではなかった。

【0005】

また、ユーザの発話に対してシステム側の応答が予め対応づけられているため、決められた言葉に対して決まりきった発話をするしかできず、自然な対話を行うという観点からも十分と言えるものではなかった。

本発明は、こうした問題に鑑みなされたものであり、ユーザとの対話状況に応じて発話内容を臨機応変に変えることができると共に、ユーザの知的好奇心にも応えることができ、知的で自然な音声対話を実現する音声対話装置を提供することを目的とする。

【0006】

【課題を解決するための手段】

上記課題に鑑み、請求項 1 記載の音声対話装置においては、ユーザから対話のための音声入力となされると、認識手段がこの入力内容を音声認識する。記憶手段には、ユーザとの対話内容に応じた複数のシナリオと、各シナリオに沿った発話対象語が予め記憶されており、選択手段が、認識手段による認識に応じて記憶手段に記憶された発話対象語の中からユーザに向けた発話語を選択し、出力手段が、この選択手段によって選択された発話語を音声により出力することにより、ユーザとの間で対話を行う。

【0007】

そして特に、学習手段が、記憶手段に記憶された発話対象語の中に、ユーザとの対話内容に応じた発話対象語がない場合に、選択手段にユーザにこの対話内容の答えを問い返すための発話語を選択させ、この問い返しに対してユーザから入力された対話内容の答えに基づき、この対話内容に応じたシナリオを学習し、このシナリオと各シナリオに沿った発話対象語を記憶手段に新たに記憶させる。

【0008】

すなわち、かかる音声認識装置においては、ユーザから知らないことを聞かれたら、その場では分からないと答えるが、次に同じようなことを聞かれたら、答えられるような学習機能を有する。つまり、ユーザから知らないことを聞かれた場合にユーザにその答えを問い返し、その質問内容と答えを記憶して、次からの対話に用いるようにする。

【 0 0 0 9 】

このため、知らない対話内容によって対話を中断させたり、ユーザの提示する話題を変更したりする必要性が小さくなると共に、学習によって新たなシナリオや語彙を増やして知識を向上させ、次回からのユーザとの対話に反映することができる。その結果、学習を重ねる毎に特定のユーザに対して満足のできる対話を実現することができるようになる。また、異なるユーザに対しては新たな話題や情報を提供することができ、知的な対話を実現することができる。

【 0 0 1 0 】

また、学習によって保有するシナリオや発話語彙のバリエーションを増加させることができ、同じ発話内容であってもその発話を様々なタイプのユーザに応じて適宜変化させることができる。このため、様々なタイプのユーザとの間で自然な対話を実現することができる。

【 0 0 1 1 】

具体的には、請求項 2 に記載のように、更新手段が、学習手段により学習された対話情報に基づき、記憶手段において、この対話内容についての音声対話に必要なシナリオ、対話辞書、認識辞書を、自動的に更新するようにすることで、これを実現することができる。

【 0 0 1 2 】

すなわち、かかる音声対話装置においては、ユーザの発話を認識するための語彙が参照される認識辞書、ユーザとの対話内容に沿った発話を実現するために予め用意された複数種類のシナリオ、各シナリオに沿った発話を実現するための語彙が参照される対話辞書が設けられている。そして、ユーザとの対話を通じて学習手段により新たに学習されたシナリオや語彙を自動的に更新し、次回から参照可能なシナリオや語彙を増加させることにより、上記知的で自然な対話を実現す

るのである。

【0013】

しかし、ユーザにより教えられたことが間違っている場合もあるので、対話を行っていく中で、1つの質問に対する答えがいくつか返ってくる場合がある。その場合に上記更新手段が対話内容を逐次更新していくと、次回からの対話においてユーザに間違った情報を提供する虞がある。

【0014】

そこで、請求項3に記載のように、第2の記憶手段が、学習手段により学習された対話情報について、複数回交わされた同種の対話内容の履歴を記憶するようにし、上記更新手段が、同種の対話内容について互いに不整合なシナリオがある場合には、対話確率の高いシナリオに順次変更して更新するようにするとよい。

【0015】

すなわち、かかる構成では、例えば異なる複数のユーザとの間で交わされた対話内容を通じて学習手段が学習した問いかけと答えとを履歴情報として記憶する。そして、その問いかけと答えの対応が複数の対話間で異なる場合には、更新手段が、その問いかけに対して最も対話確率（頻度）の高い答えを採用するようにシナリオを順次変更していく。この結果、次回からの対話においては、同じ内容の問いかけに対して、この最も頻度の高いシナリオに沿った発話をするようになる。その結果、発話内容が自然に実際の答えに近づいていくようになり、知的な対話を実現することができるのである。

【0016】

その際、対話内容についての対話確率が等しい場合も考えられるので、請求項4に記載のように、更新手段は、不整合なシナリオ間で、その対話内容についての対話確率が等しい場合には、先に出現したものを優先するようにしてもよい。

すなわち、対話確率が等しいからといって発話内容が直前に出現したものに度々されると、ユーザから優柔不断と思われ、不快感を感じられる場合がある。そこで、このように先に（より過去に）出現したものを優先することで、音声対話装置としての意志を強調して対話に勢いや信頼性を持たせることができる。

【0017】

また、請求項 5 記載の音声対話装置においては、ユーザから対話のための音声入力となされると、認識手段がこの入力内容を音声認識する。記憶手段には、ユーザとの対話内容に応じた複数のシナリオと、各シナリオに沿った発話対象語が予め記憶されており、選択手段が、認識手段による認識に応じて記憶手段に記憶された発話対象語の中からユーザに向けた発話語を選択し、出力手段が、この選択手段によって選択された発話語を音声により出力することにより、ユーザとの間で対話を行う。

【0018】

そして特に、記憶手段が、同じ意味内容の対話について複数のバリエーションのシナリオ及び対話対象語を記憶し、選択手段が、ユーザとの音声対話におけるユーザの応答内容に応じて、記憶手段から選択する発話語を変化させる。

ここでいう「ユーザの応答内容」とは、例えば後述する実施例にて説明するようなユーザの応答速度（タイミング）、ユーザの答え方、ユーザの発話内容等が該当する。

【0019】

すなわち、かかる構成では、ユーザの受け答えのタイミングや発話内容等に応じて、音声対話装置側もその応答内容やタイミングを様々に変化させるのである。例えばユーザの受け答えが早い場合には、その事柄について興味がある、又はよく知っていることである可能性が高いので、装置側の応答も間をあけず、その事柄を強調するような発話をし、その話題が続くようなら、深く対話を進めるようにすることが考えられる。逆に、ユーザの受け答えが遅ければ、その事柄について興味がない、又は答えが曖昧であるという可能性が高いので、装置側の応答も、曖昧性を持たせた発話を返すようにすることが考えられる。さらに、タイミングだけでなく、ユーザの発話内容、例えば、発話文章の語尾の違い（「・・だよ」＜断定＞、「・・かな」＜曖昧＞、等）によっても、装置側がユーザの発話内容に合わせて発話を強調したり、曖昧性を持たせたりすることもできる。このような装置側の発話の変化により、自然な音声対話を実現することができる。

【0020】

さらに、請求項 6 に記載のように、識別手段が、ユーザからの音声入力に基づ

いてユーザの感情を識別し、選択手段が、この識別手段により識別されたユーザの感情情報に応じて、出力手段が出力する発話語の語調を変化させるように、記憶手段から発話語を選択するようにしてもよい。

【 0 0 2 1 】

ここでいう「ユーザの感情」（喜怒哀楽）は、例えばユーザの発話音声の速さ、高さ、大きさ、発話語自体等から判断される口調により識別される。そして、例えばユーザが怒っているような口調で話し掛けた場合には、なだめるようなやさしい言葉で発話したり、ユーザが喜んでいる場合には、テンションを上げて気分をさらに高揚させるような発話をしたりすることにより、ユーザとの間でその後のより自然で円滑な対話を実現することができる。

【 0 0 2 2 】

その際、請求項 7 に記載のように、識別手段が、ユーザの感情情報が当該音声対話装置に対してのものか、又は一般的なことに対してのものかを識別し、選択手段が、この識別手段による識別結果に応じて、出力手段が出力する発話語の語調を変化させるように、記憶手段から発話語を選択するようにするのがよい。

【 0 0 2 3 】

すなわち、ユーザの感情が変わったとしても、その原因（喜怒哀楽の対象）が当該音声対話装置側の発言によるものなのか、対話内容に現れる一般的事象についてのことなのかによって、ユーザをなだめたり、同調したりする等の対応を変えるのである。かかる構成により、より人間の対話に近い知的で自然な対話を実現することができる。

【 0 0 2 4 】

さらに、請求項 8 に記載のように、識別手段が、さらにユーザからの音声入力に基づいて方言を識別し、選択手段が、この識別手段により識別された方言に応じて、出力手段が出力する発話語の語調を変化させるように、記憶手段から発話語を選択するようにしてもよい。

【 0 0 2 5 】

かかる構成によれば、ユーザの方言と同じ方言で対話することにより、ユーザに対して親しみを持たせたり、逆にユーザの方言と異なる方言で対話することに

より、対話に面白みを持たせたりすることができる。

前者の場合には、請求項 9 に記載のように、選択手段が、識別手段により識別された方言に応じて、対話の話題がこの方言にかかる土地柄にちなんだものになるように、記憶手段から発話語を選択可能に構成されていてもよい。

【0026】

すなわち、対話における話題の転換に際して、ユーザの方言を手がかりにしてユーザにとって親しみ深い又は知識の豊富な話題に転換させることにより、ユーザが当該音声認識装置との対話に積極的になることができ、対話を自然に一層楽しむことができる。

【0027】

或いは、請求項 10 に記載のように、識別手段が、さらにユーザからの音声入力に基づいてその言語を識別し、選択手段が、この識別手段により識別された言語に応じて、出力手段が出力する発話語の語調を変化させるように、記憶手段から発話語を選択するようにしてもよい。

【0028】

かかる構成によれば、ユーザの言語と同じ言語で対話することにより、ユーザの理解が容易になり、国籍に拘わらず自然で円滑な対話を実現することができる。

この場合にも、請求項 11 に記載のように、選択手段が、識別手段により識別された言語に応じて、対話の話題がこの言語にかかる国にちなんだものになるように、記憶手段から発話語を選択可能に構成されていてもよい。

【0029】

かかる構成により、異国籍のユーザにとって親しみ深い又は知識の豊富な話題に転換させることにより、ユーザが当該音声認識装置との対話に積極的になることができ、対話を一層楽しむことができる。特に母国を離れたユーザにとっては、懐かしみや安堵感を与えることができる。

【0030】

また、請求項 12 に記載のように、判定手段が、ユーザからの音声入力に基づいてその声質からユーザの属性を判定し、選択手段が、判定手段により判定され

た属性に応じて、出力手段が出力する発話語の声質を変化させるように、記憶手段から発話語を選択するようにしてもよい。

【 0 0 3 1 】

かかる構成では、ユーザから入力された音声の高さ、太さ、大きさ、発話の仕方等の声質からユーザの年齢や性別等の属性を判定し、その属性と対話状況等に応じて適切な声質で応答する。例えば、小さい子供に対しては、幼稚園の先生のようなお姉さんの声で対応し、男の人には女の人で、女の人には男の人で応答することが考えられる。かかる構成により、ユーザに対話への欲求を高めさせたり、対話をより楽しませることができる。

【 0 0 3 2 】

或いは、請求項 1 3 に記載のように、判定手段が、ユーザの姿態を撮像して画像認識してユーザの属性を判定し、選択手段が、判定手段により判定された属性に応じて、出力手段が出力する発話語の声質を変化させるように、記憶手段から発話語を選択するようにしてもよい。

【 0 0 3 3 】

かかる構成により、請求項 1 2 と同様の効果を得ることができるが、画像認識によりユーザの属性を判定するため、その属性の判定結果がより正確となる可能性が高くなる。

その際、請求項 1 4 に記載のように、選択手段が、判定手段により判定された属性に応じて、対話の話題がこの属性にちなんだものになるように、記憶手段から発話語を選択可能に構成されたものでもよい。

【 0 0 3 4 】

かかる構成により、対話における話題の転換に際して、ユーザにとって興味深い、親しみ深い又は知識の豊富な話題に転換させることにより、ユーザが当該音声認識装置との対話に積極的になることができ、対話を自然に一層楽しむことができる。

【 0 0 3 5 】

また、請求項 5 に記載の音声対話装置においては、ユーザから対話のための音声入力がなされると、認識手段がこの入力内容を音声認識する。記憶手段には、ユ

ーザとの対話内容に応じた複数のシナリオと、各シナリオに沿った発話対象語が予め記憶されており、選択手段が、認識手段による認識に応じて記憶手段に記憶された発話対象語の中からユーザに向けた発話語を選択し、出力手段が、この選択手段によって選択された発話語を音声により出力することにより、ユーザとの間で対話を行う。

【 0 0 3 6 】

そして特に、画像認識手段が、ユーザの顔画像を撮像し、その唇の動きに基づくリップリーディングにより画像認識し、認識手段が、画像認識手段による画像認識を併用して音声認識を行う。

かかる構成では、ユーザから音声入力された発話語の認識に際し、唇の動きを解析してユーザの発話語を解析する所謂リップリーディングによる画像認識が併用される。例えば、認識対象となる発話語の正確な認識率が、音声認識による方が高い発話語、リップリーディングによる画像認識による方が高い発話語、音声認識及び画像認識の双方によるマッチングによるのが良い発話語等、発話語の種類等によって認識方法を予めデータベース化しておき、それにより判定するようにすることができる。

【 0 0 3 7 】

このように画像認識を併用することで、ユーザの発話語以外のノイズを除去して音声認識することができ、認識手段による発話語の認識率が向上する。それにより、ユーザの発話に対する装置側の錯誤が防止又は抑制することができ、ユーザとの間で自然な対話を実現することができる。その結果、ユーザとの間で知的な対話を進めることができる。

【 0 0 3 8 】

尚、以上に述べた音声対話装置は、請求項 1 6 に記載のように、ユーザと対話するロボットとして構成することができる。

つまり、音声対話装置を人間の姿態に近似したロボットとして構成することにより、人間間の対話を擬似することができ、ユーザにとってより自然な対話を実現することができる。

【 0 0 3 9 】

この場合、請求項 1 7 に記載のように、これをユーザの顔画像を撮像する目を備えたロボットとして構成し、画像認識手段が、この目により撮像された顔画像から、ユーザがロボットの正面を向いているか否かを判定し、認識手段が、画像認識手段によりユーザがロボットの正面を向いていると判定された場合にのみ、音声認識を行うようにすることが考えられる。

【 0 0 4 0 】

このように、ロボットにユーザの顔が正面を向いているときの音声のみを認識するようにさせることで、ノイズ対策が行え、認識率の向上につながる。

その際、請求項 1 8 に記載のように、ロボットの目が四方を見渡せるように、その頭部周囲に複数設けられ、画像認識手段が、この複数の目のいずれかにより撮像されたユーザの顔画像により、ユーザがロボットの方向を向いているか否かを判定し、認識手段が、画像認識手段によりユーザがロボットの方向を向いていると判定された場合にのみ、音声認識を行うようにしてもよい。

【 0 0 4 1 】

かかる構成によれば、ロボットがその目（「カメラ」等）により四方（3 6 0 度）を見渡せるため、ロボットの後方からの音声であっても、ユーザがロボットの方向を向いて話し掛けてきた音声を特定して認識することができる。また、ロボットに話しかけられてない全く関係のない音声（ノイズ）についてはその認識をしないことで、ロボットにかかる処理負担を軽減する一方で、ユーザにとっては、自己が話しかけないロボットが突然対話に介入して驚かされることもなく、自然な対話を実現することができる。また、ロボット自身は、このようなユーザの顔の位置認識により、本当に認識したい語彙のみ認識する知的な音声対話ロボットとなる。ただし、ここでいうロボットの「目」は、必ずしもユーザからその全てをロボットの目として認識できるものである必要はなく、個々の目に撮像できる機能が備わっていればよい。つまり、複数の目のいずれか 2 つが、ユーザからロボットの目として認識できるように構成されていたほうが、ユーザがロボットの顔を人間と同様に認識できて好ましいとも考えられる。

【 0 0 4 2 】

また、請求項 1 9 に記載のように、認定手段を構成するロボットの耳がその頭

部周囲に複数設けられ、認識手段が、この複数の耳に入力される音声レベルに基づいてユーザがロボットの方向を向いているか否かを判定し、ユーザがロボットの方向を向いていると判定された場合にのみ、音声認識を行うようにしてもよい。

【0043】

かかる構成において、このロボットの耳は、例えば複数の指向性マイク等により構成され、ユーザの発話により入力される音声レベルの大きさやその音声レベルの変化により、ユーザがロボットの方向を向いて話し掛けてきたかどうか、また、どの方向から話し掛けてきたか等を認識することができる。このため、ロボットの後ろの方からの音声であっても、ユーザがロボットの方向を向いて話し掛けてきた音声を特定して認識することができる。その結果、請求項18に記載の効果と同様の効果を得ることができる。

【0044】

さらに、請求項20に記載のように、画像認識手段によりユーザとロボットが向き合っていないと判定された場合に、ロボットがユーザに向き合うようにロボットを動作させるようにしてもよい。これにより、人間が行っている会話のように、自然な動作や対話となる。

【0045】

尚、このような音声対話装置の各手段をコンピュータにて実現する機能は、例えば、コンピュータ側で起動するプログラムとして備えることができる（請求項21）。このようなプログラムの場合、例えば、FD、MO、DVD、CD-ROM、ハードディスク等のコンピュータ読取可能な記録媒体に記録し、必要に応じてコンピュータにロードして起動することにより用いることができる。その他、ROMやバックアップRAMをコンピュータ読取可能な記録媒体としてプログラムを記録しておき、このROM或いはバックアップRAMをコンピュータに組み込んでよい。尚、ここでいう「各手段」とは、各請求項中の各構成要件としての個々の手段を意味するのではなく、請求項単位の手段の集まりを意味する。

【0046】

【発明の実施の形態】

以下、本発明の実施の形態を具体化した実施例を図面と共に説明する。図 1 は本実施例の音声対話装置の全体構成を表すブロック図である。

1. 音声対話装置の構成

同図に示すように、音声対話装置 1 は、音声対話ロボットとして構成され、音声認識部 1 0、シナリオインタープリタ 2 0、対話シナリオ部 3 0、顔画像認識判定部 4 0、ロボット発話語決定部 5 0、学習機能部 6 0、及び音声合成部 7 0等を備えている。また、ユーザの姿態を撮像可能なカメラがその頭部の周りに一定間隔で複数設けられており、ユーザがたとえロボットの後方から話し掛けてきても、これを認識することができるようになっている。そして、ユーザからはその複数のカメラのうちの 2 つがロボットの目として認識できるように構成されている。さらに、ユーザの音声を入力するための指向性マイクが、その頭部の周りに一定間隔で複数設けられており、四方からユーザの発話音声を入力できるようになっている。そして、ユーザからはその複数の指向性マイクのうちの 2 つがロボットの耳として認識できるように構成されている。

【0047】

そして、ユーザの発話音声は、上記指向性マイクを介してまず音声認識部 1 0に入力される。音声認識部 1 0 は、ユーザの発話により指向性マイクから入力される音声レベルの大きさやその音声レベルの変化等により、ユーザがロボットの方向を向いて話し掛けてきたかどうか、また、どの方向から話し掛けてきたか等を認識することができる。また、顔画像認識判定部 4 0 は、これと同時に上記複数のカメラから入力されたユーザの顔画像からその顔の位置や向きを判定し、ユーザがロボットの方向を向いて話し掛けてきたかどうかの判定精度を向上させたり、ユーザの唇の動きを解析して所謂リップリーディングによる画像認識を行い、音声認識の精度を向上させることができる。

【0048】

そして、音声認識部 1 0 は、話し掛けてきたユーザの発話音声を認識すると、対話に必要な語彙が格納された認識辞書 1 1 を参照してこの発話音声の内容を認識し、この認識結果をシナリオインタープリタ 2 0 に出力する。

対話シナリオ部 3 0 には、対話上の条件分岐等を表す複数種類のシナリオが格

納されている。この対話シナリオ部 3 0 は、シナリオインタープリタ 2 0 を介して得た上記認識結果、時間計測器 8 0 による経過時間情報等を参照して、対話の進行状況に適合したシナリオ（発話語）を生成し、その情報をシナリオインタープリタ 2 0 に出力する。

【 0 0 4 9 】

シナリオインタープリタ 2 0 は、対話シナリオ部 3 0 にて決定されたシナリオに従って、対話用認識辞書 2 1 及び発話リスト格納部 2 2 を参照し、次の発話内容を設定するための演算処理を行う。ここで、対話用認識辞書 2 1 には、対話用の単語等の装置で用いられる単語が格納され、発話リスト格納部 2 2 には、対話のシナリオに応じて複数設定された文章化された発話語が選択可能に格納されている。

【 0 0 5 0 】

さらに、シナリオインタープリタ 2 0 が、発話リスト格納部 2 2 や対話用認識辞書 2 1 を参照しても、ユーザとの対話内容に応じた発話対象語がない場合には、ユーザにこの対話内容の答えを問い返すことになる。その際、学習機能部 6 0 が、この問い返しに対してユーザから入力された対話内容の答えに基づき、この対話内容に応じたシナリオとこのシナリオに沿った発話対象語（語彙）を学習し、当該シナリオ及び発話対象語をシナリオインタープリタ 2 0 を介して、対話用認識辞書 2 1，発話リスト格納部 2 2，対話シナリオ部 3 0 に格納して更新し、次回からの対話に反映させる。つまり、ここでは円滑でかつ適切な対話をするために、シナリオの作成、及びそれに伴う対話辞書、発話リストの作成が行われる。

【 0 0 5 1 】

そして、ユーザとの発話において、ロボット発話語決定部 5 0 が、対話シナリオ部 3 0 に新たに格納されたシナリオ等も含めてロボットの発話語を決定し、これに対応した発話語を発話リスト格納部 2 2 からシナリオインタープリタ 2 0 に出力させる。そして、シナリオインタープリタ 2 0 にて最終的に生成された応答内容が、音声合成部 7 0 にて音声合成され、ロボットの発話としてスピーカから出力される。

2. 学習機能（知的な対話）

音声対話を進めていくうちに、ロボットが答えられない（シナリオに記述されていない）ことをユーザから聞かれることが出てくる。その際、はじめは分からないので、ユーザにその答えを問い返す。このことにより、ロボットはその質問内容と答えを学習し、自動的に音声対話装置 1 に必要なシナリオ、対話辞書、認識辞書を更新する。したがって、2 回目以降は、今まで答えられなかったことに対して、答えられるようになっていく。しかし、学習した答えが間違っている場合もあり、対話を行っていく中で、1 つの質問に対して、いくつかの答えが発生する場合が出てくる。その場合は、その答えの中での出現確率が一番高いものをロボットが発話する答えとする。等確率のものが発生した場合は、先に出現したものを優先する。このような機能を備えることで、知らないことを学習できるようになり、また、その答えも高い確率で正確な答えに近づいていくようになる。

3. ロボット発話（自然な対話）

ロボットが応答する発話を、質問に対して、毎回同じことを発話するのではなく、ユーザの応答時間間隔や答え方、発話内容などによって、ロボットの発話も様々に変化させることができる。また、ユーザの感情や方言など、様々な要因によってもロボット発話を変化させることができる。

4. 作動

次に、図 2 ～図 4 に示すフローチャートに基づいて、本実施例の音声対話装置の動作について説明する。

4. 1 全体の流れ

本実施例の音声対話装置 1 の全体の流れとしては、対話シナリオ部 3 0 で設定したシナリオどおりに進んでいく。ユーザの応答待ち、すなわちシナリオの各分岐点において、図 3 に示す学習機能に関する動作フロー、図 4 に示すロボット発話に関する動作フローを適用し、それに対応するロボットの発話を出力していく

。この操作を繰り返し、シナリオにより、対話終了となった時点で終了とする。

【 0 0 5 2 】

すなわち、図 2 に示すように、ユーザの応答を待ち（S 1 1 0）、まず顔画像認識判定処理を実行する（S 1 2 0）。

この顔画像認識判定処理においては、上述したロボットの複数の目（カメラ）及び耳（マイク）を介して、音声認識部 1 0 が、ユーザがロボットの方向を向いて話し掛けてきたか否かを判定し、図 1 0 に示すように、ユーザがロボットの方向を向いて話し掛けてきたと判定されると音声認識を開始する。そのとき、発話中にユーザが顔を背けても、認識をストップさせず、その発話が完了するまで認識は止めない。また、ユーザがロボットの方向を向いていなくて入ってきた音声は認識せず、その途中で、ユーザがロボットの方向に顔を向けた場合は、その時点から認識を開始させるようにする。この場合、上述のように、ロボットにはその頭部周囲の 3 6 0 度全てにわたって一定の間隔で、目（カメラ）及び耳（マイク）が設けられているため、ロボットの後ろの方からの音声でも、ユーザがロボットの方向を向いて話し掛けてきた音声は認識することができる。本実施例では、その際、ロボットがそのユーザの方向を向くようにされており、これにより、人間が行っている会話のように自然な動作／対話をするようになっている。こうすることで、ロボットは、本当に認識したい語彙のみ認識するようになる。

【 0 0 5 3 】

そして、ユーザの顔画像を認識したと判定されると（S 1 3 0 : Y E S）、続いて、図 3 に示す学習機能に関する動作フローを実行する（S 1 4 0）。

4. 2 学習機能（知的な対話）

音声対話を進めていくうちに、ロボットが答えられない（シナリオに記述されていない）ことをユーザから聞かれることが出てくる。本学習機能は、このような事態に対応できる知的な対話を実現するものである。尚、図 5 及び図 6 には、ユーザに知らないことを聞かれたときの対応例が示されている。

【 0 0 5 4 】

すなわち、ユーザとの対話を通じて、シナリオインタープリタ 2 0 が対話シナ

リオ部 3 0 を参照し、ユーザからの発話内容（質問内容）についてロボットが知っている内容であるか否かを判定する（S 2 1 0）。このとき、対応するシナリオがなく、知らない内容であると判定されると（S 2 1 0 : N O）、ロボット側からユーザに対して、「分からないので、教えて」などと言ってその答えを問い返し（S 2 2 0）、これに対するユーザの回答に基づき、学習機能部 6 0 がその質問内容とその答えを学習する（S 2 3 0）。そして、このとき得られた新たなシナリオや発話語彙をシナリオインタープリタ 2 0 を介して対話用認識辞書 2 1、発話リスト格納部 2 2、及び対話シナリオ部 3 0 等に格納する。つまり、全く知らないことに対してはシナリオを増やしていき、言葉の意味が分からないだけの場合には、認識語彙を増やしていく。このことにより、2 回目以降は、分からなかったことに対しても答えられるようになり、新たなシナリオや発話語彙が蓄積されていくことで知的なロボットとなっていく。具体例を示すと、図 5 に示す如くである。

【 0 0 5 5 】

一方、S 2 1 0 において、対応するシナリオがあり、知っている内容であると判定されると（S 2 1 0 : Y E S）、そのシナリオに記述されている発話を選択し、音声合成部 7 0 にて音声合成して発話する（S 2 4 0）。そして、この発話に対してユーザに誤りを指摘されなければ（S 2 5 0 : N O）、当該動作フローを終了する。

【 0 0 5 6 】

ただし、はじめに答えてもらったユーザの答えが必ずしも正解であるとは限らない。このため、2 回目以降にユーザ側から「違うよ」と指摘されることも想定される。このため、S 2 5 0 において、ユーザに誤りを指摘された場合には（S 2 5 0 : Y E S）、学習機能部 6 0 が、まずその質問内容とその答えを学習する（S 2 6 0）。そして、その質問内容と同じ対話内容について、発話リスト格納部 2 2 から過去の履歴を参照し、その質問内容に対して現在一番出現確率（累積値）の高い答えとの確率の比較を行う（S 2 7 0）。

【 0 0 5 7 】

このとき、今回の答えの出現確率が大きいと判定されると、当該質問内容につ

いての答えを正解とみなし、次回からの答えに変更し、シナリオを更新して当該動作フローを終了する（S280）。一方、今回の答えの出現確率が小さいと判定されると、当該質問内容についての答えの変更は行わない（S290）。さらに、両者の確率が等しい場合には、先に（より過去に）出現したものを、次回の発話に使用するように設定する（S300）。これらの具体例を示すと、図6に示す如くである。

【0058】

図2に戻り、続いて図4に示すロボット発話に関する動作フローを実行する（S150）。

4. 3 ロボット発話（自然な対話）

本動作フローでは、図4に示すように、ロボット側の発話に対するユーザの応答時間間隔（S310）、ユーザの答え方（S340）、及びユーザの発話内容の判断（S350）に基づき、その発話内容を強調した発話をしたり（S320）、曖昧性を持たせた発話をしたりする（S330）。

【0059】

具体的には図7に示すように、例えばロボットがユーザに対して、「好きな食べ物」と聞いた場合、ユーザがすぐに、例えば「リンゴだよ」と答えたとすると、ロボットは「そうですか。リンゴが大好きなんだね。」などと、好きなことを強調するような発話にする。逆に、ユーザが間を開けて（例えば10秒程度）、「リンゴかな」と答えたとすると、ロボットは「本当にリンゴが好きなの。」といったような、ユーザが考えて出した答えに対して、曖昧性を持たせた返答とする。このように、返答に差をつけることで、意味、感情といった部分を考慮に入れた知的なロボットとなる。

【0060】

また、ユーザの答え方、例えば、「リンゴだよ」と「リンゴかな」の違いのように、「だよ」とあると確信を持った断定的な言い方であるし、「かな」とあると少し曖昧性を持った言い方であるので、このような点を見極めて、ロボット発話の返答をかえる。

【 0 0 6 1 】

さらに、ユーザの発話内容から、「えーと、・・・」などと、頭に語彙が入ると、考えていて、あまり確信がなく曖昧な言い方と受け取れるので、このような点も考慮に入れて、ロボット発話の返答をかえる。

図2に戻り、以上のようにして決定されたロボットの発話内容に従って発話を行い（S 1 6 0）、続いて終了条件判定処理を実行する（S 1 7 0）。そして、シナリオに基づいて予め設定した終了条件を具備したと判定されると（S 1 8 0 : Y E S）、一連の処理を終了する。

5. その他の知的で自然なロボット音声対話に関する要因

図8に知的で自然なロボット音声対話装置に関する要因を示す。

5. 1 感情

図8に示すように、ユーザの発話に含まれる感情を認識し、ロボットの発話口調を変化させる。例えば、ユーザがロボットに対して、怒っているような口調で話し掛けた場合には、ロボットの発話は、なだめるようなやさしい言葉で発話させるようにする。

【 0 0 6 2 】

また図9に示すように、そのユーザの感情の対象の違い、例えば、ロボットの発話に対しての感情か、一般的なことに対しての感情かによっても、ロボット発話を変化させることができるようにする。

5. 2 方言

図8に示すように、ユーザの発話の方言に対して、ロボットの発話も同様の方言を用いて発話させることで、親しみのわく知的なロボットとする。例えば、ユーザが関西弁で話し掛けた場合には、ロボットの発話も関西弁にするといった具合である。

【 0 0 6 3 】

また、方言の認識より、話題をその土地柄にちなんだものに進めていくように

する。こうすることで、話題の転換ができ、ユーザにとっても答えやすい話題へと進んでいくという工夫を入れている。

5. 3 言語

図 8 に示すように、ユーザの発話の言語に対して、ロボットの発話も同様の言語で発話させる。例えば、英語で話し掛けられたら、ロボットも英語で発話するといった具合である。

【0064】

また、言語の認識より、話題をその国にちなんだものに進めていくようにする。こうすることで、話題の転換ができ、ユーザにとっても答えやすい話題へと進んでいくという工夫を入れている。

5. 4 発話音声

図 8 に示すように、ユーザの年齢や性別に応じて、ロボットの発話音声を変化させる。例えば、小さい子供に対しては、幼稚園の先生のようなお姉さんの声で対応し、男の人には女の人で、女の人には男の人の声で応答するといった具合である。

【0065】

また、発話音声だけでなく、年齢や性別にちなんだ話題へと進めていくことができるという工夫を入れている。

5. 5 画像（目をもたせる）

図 8 に示すように、1 つには、目をもたせることで、ユーザの唇の動きを見ることができるので、リップリーディングの技術を応用し、認識率の向上に努めることができ、どんな言葉でも認識できるという知的なロボットに役立つ。

【0066】

もう 1 つには、ユーザの顔の位置を認識し、ロボットはユーザの顔が正面を向いていると判定したときの音声のみを認識することで、ノイズ対策が行え、認識率向上につながる。

具体的には、上記において図 1 0 に基づいて説明したとおりである。

6. 対話例

図 1 1 にロボット音声対話の対話例を示す。

:

:

ロボット 「A さん、好きな食べ物は」

ユーザ A 「リンゴだよ」 ← (すぐに答えた)

ロボット 「A さんはリンゴが大好きなんだね」

ロボット 「じゃあ、B さんの好きな食べ物は」

ユーザ B 「えーとね、バナナかな」 ← (1 0 秒程度考えてから答えた)

ロボット 「B さん、本当にバナナが好きなのかな」

:

:

ユーザ A 「ロボットさん、サッカーは何人でやるか知ってる」

← (知らないことを聞かれた)

ロボット 「分かんないな。教えてよ」 ← (ユーザに問い返す)

ユーザ A 「サッカーはね、1 1 人でやるんだよ、ロボットさん」

ロボット 「教えてくれてありがとう」

← (サッカーの人数の質問と答えを学習)

:

:

ユーザ C 「ところでロボット君、サッカーって何人でやるのかな」

ロボット 「サッカーは 1 1 人でやるよ」

← (前に一度聞かれていて学習

したので答えることができた)

ユーザ C 「よく知ってるね」

:

:

以上に説明したように、本実施例の音声対話装置 1 においては、ユーザから知らないことを聞かれたら、その場では分からないと答えるが、次に同じようなことを聞かれたら、答えられるような学習機能を有する。つまり、ユーザから知らないことを聞かれた場合にユーザにその答えを問い返し、その質問内容と答えを記憶して、次からの対話に用いるようにする。

【 0 0 6 7 】

このため、知らない対話内容によって対話を中断させたり、ユーザの提示する話題を変更したりする必要性が小さくなると共に、学習によって新たなシナリオや語彙を増やして知識を向上させ、次回からのユーザとの対話に反映することができる。その結果、学習を重ねる毎に特定のユーザに対して満足のできる対話を実現することができるようになる。また、異なるユーザに対しては新たな話題や情報を提供することができ、知的な対話を実現することができる。

【 0 0 6 8 】

また、ユーザの受け答えのタイミングや発話内容等に応じて、音声対話装置 1 側もその応答内容やタイミングを様々に変化させる。このような装置側の発話の変化により、自然な音声対話を実現することができる。

さらに、音声対話装置 1 としてのロボットに複数の目（カメラ）や耳（マイク）を設けたり、リップリーディングの技術を応用することにより、ユーザの発話語以外のノイズを除去して音声認識することができ、認識手段による発話語の認識率が向上する。それにより、ユーザの発話に対する装置側の錯誤が防止又は抑制することができ、ユーザとの間で自然な対話を実現することができる。その結果、ユーザとの間で知的な対話を進めることができる。

【 0 0 6 9 】

尚、本実施例において、音声認識部 1 0 が認識手段に該当し、顔画像認識判定部 4 0 が画像認識手段に該当し、シナリオインタープリタ 2 0、対話シナリオ部 3 0 及びロボット発話語決定部 5 0 が、選択手段、判定手段、更新手段、識別手段に該当する。また、対話用認識辞書 2 1、発話リスト格納部 2 2 が、記憶手段

、第 2 の記憶手段に該当し、学習機能部 6 0 が学習手段に該当し、音声合成部 7 0 が出力手段に該当する。

【0 0 7 0】

以上、本発明の実施例について説明したが、本発明の実施の形態は、上記実施例に何ら限定されることなく、本発明の技術的範囲に属する限り種々の形態をとり得ることはいうまでもない。

例えば、上記実施例では、本発明の音声対話装置をロボットとして構成した例を示したが、これに限らず、ナビゲーションシステム等の装置として構成してもよいことはもちろんである。

【図面の簡単な説明】

【図 1】 本発明の実施例に係る音声対話装置の概略構成を表すブロック図である。

【図 2】 実施例の音声対話装置の動作を表すフローチャートである。

【図 3】 音声対話装置の学習機能の動作を表すフローチャートである。

【図 4】 音声対話装置に係るロボット発話動作を表すフローチャートである。

【図 5】 ロボットが知らないことを聞かれたとき（学習機能）の対応例 1 を表す説明図である。

【図 6】 ロボットが知らないことを聞かれたとき（学習機能）の対応例 2 を表す説明図である。

【図 7】 ユーザの受け答えのタイミングの違いによるロボットの対応例を表す説明図である。

【図 8】 知的なロボットに関する要因を表す説明図である。

【図 9】 ユーザの感情の対象の違いによるロボット発話の対応例を表す説明図である。

【図 1 0】 ロボットの画像認識による音声対話の例を表す説明図である。

【図 1 1】 ロボット音声対話の対話例を表す説明図である。

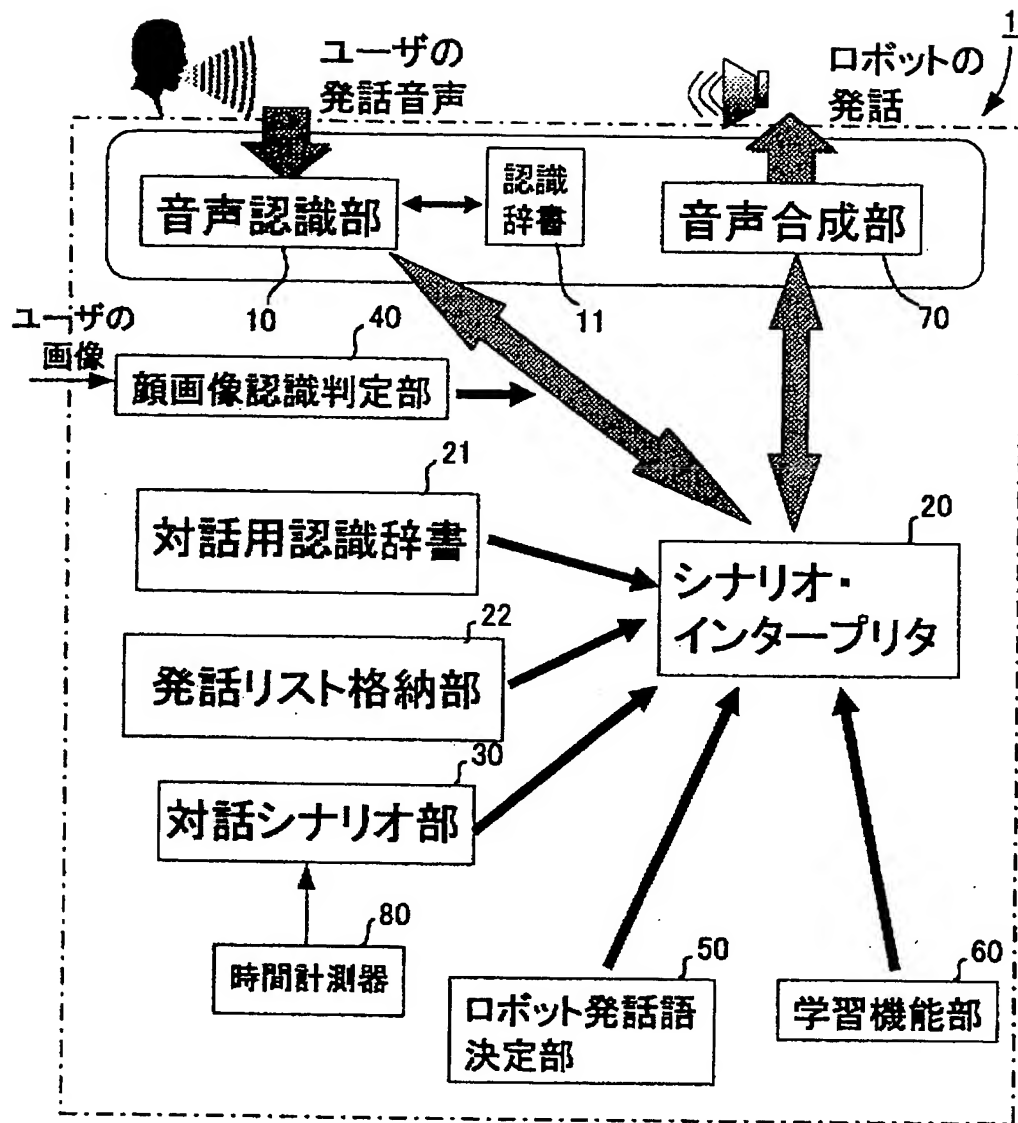
【符号の説明】

1・・・音声対話装置、 1 0・・・音声認識部、

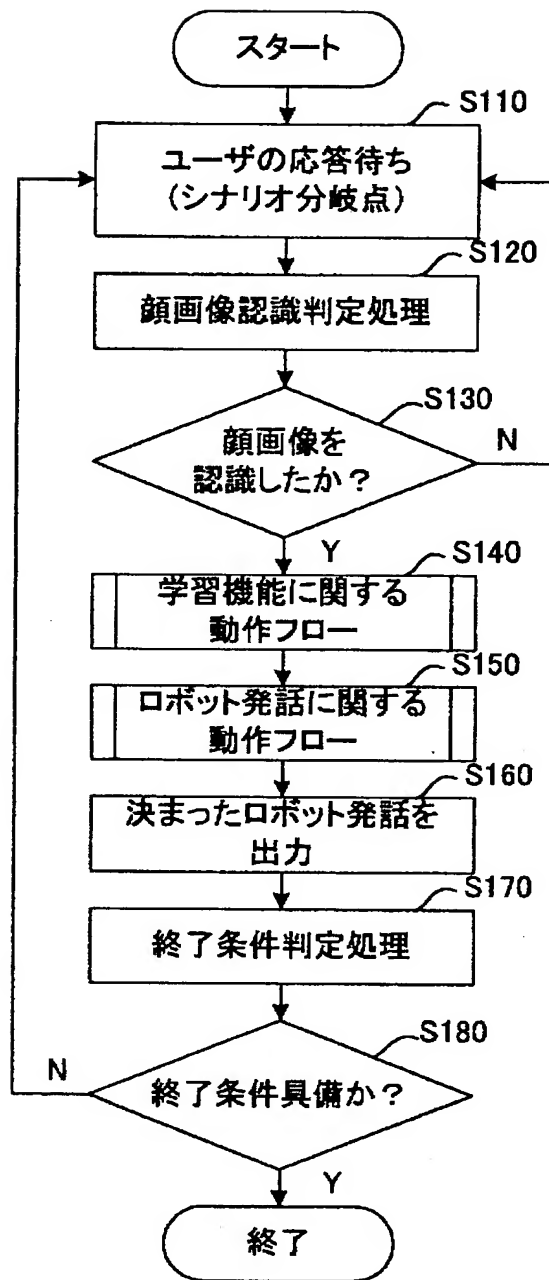
2 0 . . . シナリオインタープリタ、 2 1 . . . 対話用認識辞書、
2 2 . . . 発話リスト格納部、 3 0 . . . 対話シナリオ部、
4 0 . . . 顔画像認識判定部、 5 0 . . . ロボット発話語決定部、
6 0 . . . 学習機能部、 7 0 . . . 音声合成部

【書類名】 図面

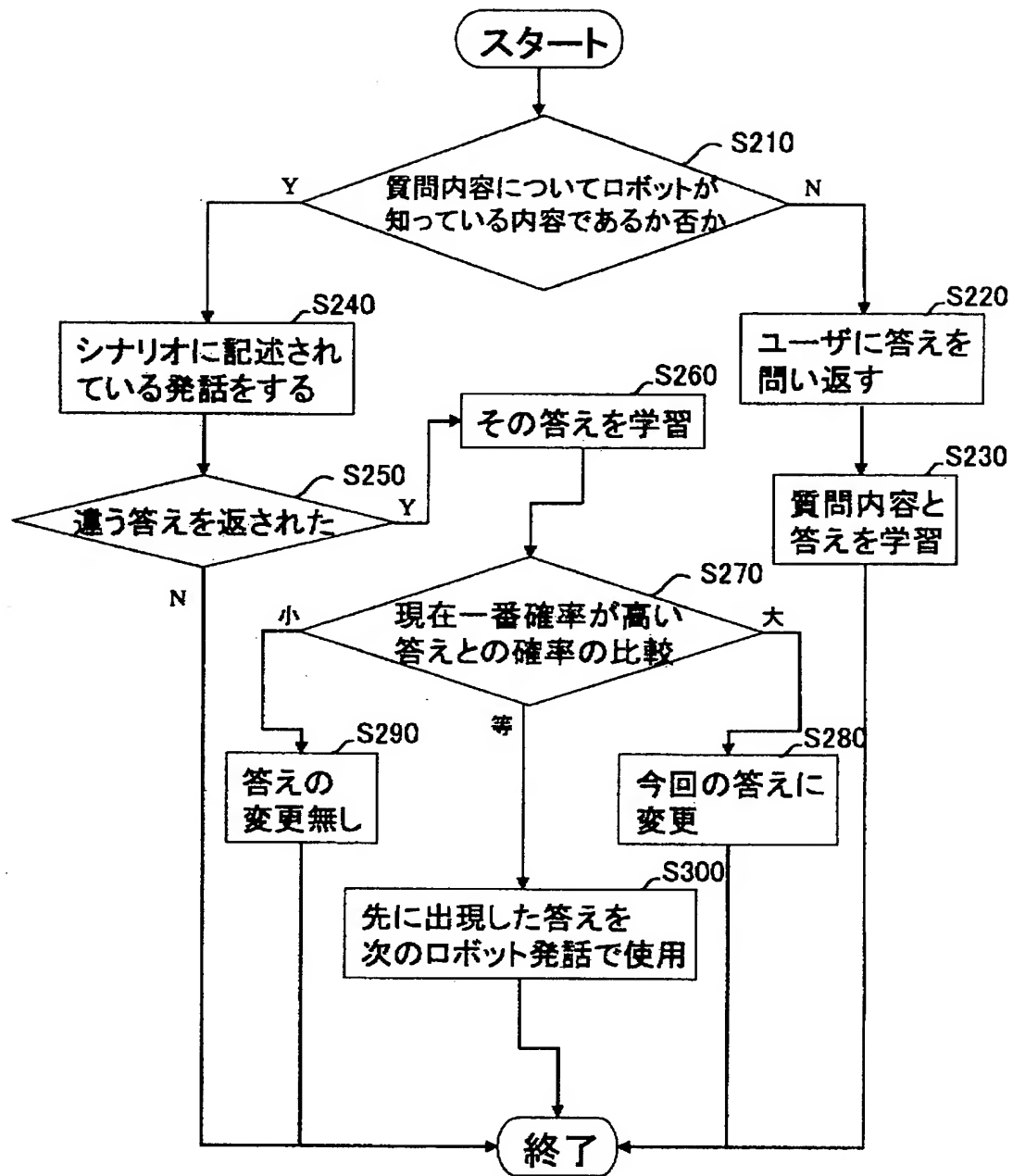
【図 1】



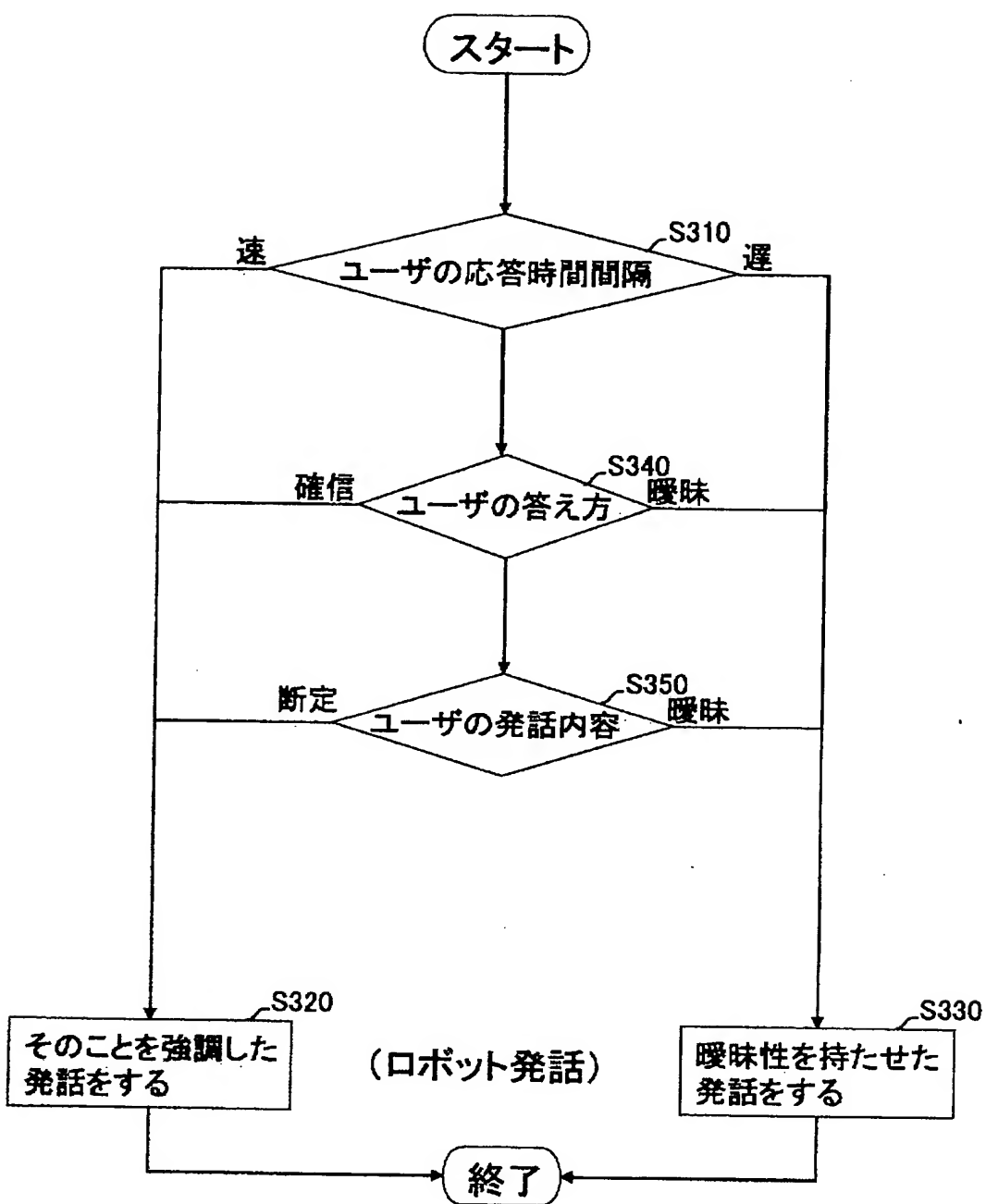
【図 2】



【図 3】



【図 4】



【図5】

(ユ)野球は何人でやるのかな。
-1回目(知らない)
 (ロ)わかんないな。教えて。
 (ユ)9人だよ。
-2回目以降(一度聞かれている)
 (ロ)野球は9人でやるよ。

【図6】

(ユ)野球は何人でやるのかな。
-1回目(知らない)
 (ロ)わかんないな。教えて。
 (ユ)8人だよ。
-2回目(一度間違って教えられた)
 (ロ)野球は8人でやるよ。
 (ユ)野球は9人だよ。
 (ロ)え、違うよ、8人だよ。
-3回目(確率50/50⇒初めの答え)
 (ロ)野球は8人でやるよ。
 (ユ)野球は9人だよ。
 (ロ)え、そうなの。
-4回目(9人の方が確率高い)
 (ロ)野球は9人でやるよ。

【図7】

(ロ)好きな食べ物は。
-ユーザがすぐに答えた場合
 (ユ)リンゴだよ。
 (ロ)そうか。リンゴが大好きなんだね。
-ユーザが少し間をあけて答えた場合
 (ユ)リンゴかな。
 (ロ)本当にリンゴが好きなの。

【図 8】

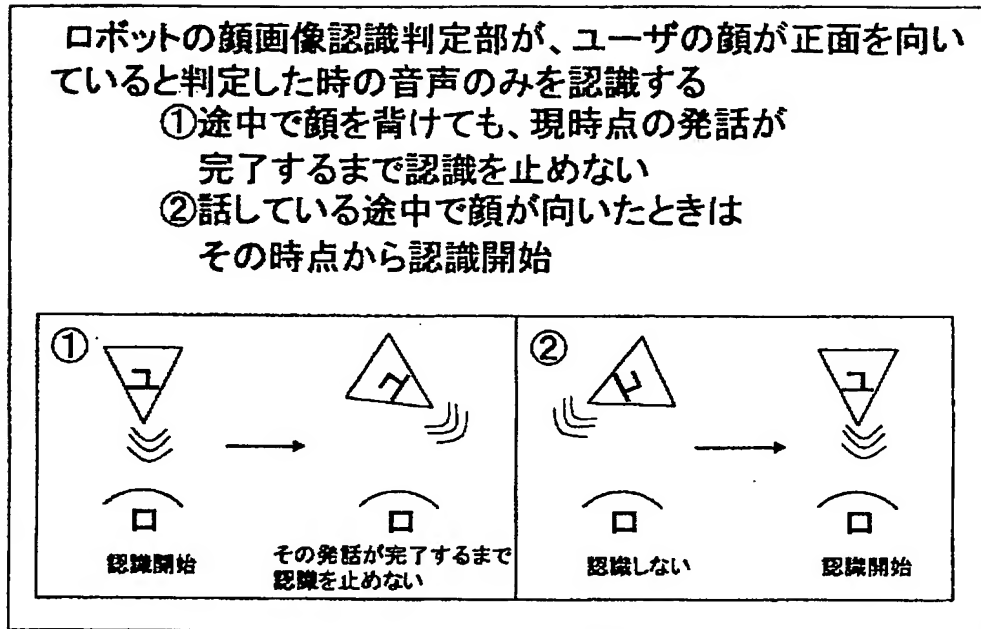
感情⇒(喜怒哀楽によって、発話の口調などを変化させる)
 方言⇒(ユーザと同様の方言で発話させる)
 ⇒(その土地柄にちなんだ話題に持っていく)
 言語⇒(ユーザと同様の言語で発話させる)
 ⇒(その国にちなんだ話題に持っていく)
 発話音⇒(年齢、性別の変化に応じた発話音声とする)
 ⇒(年齢や性別にちなんだ話題に持っていく)
 画像(目をもたせる)
 ⇒(リップリーディングによる認識率向上)
 ⇒(顔が向いているときに発話を認識するというノイズ対策)

【図 9】

●ロボットの質問に対しての怒⇒謝る
 【ロ】あなたは何歳。
 【ユ】女性に年齢は聞かないの。
 【ロ】ごめんなさい。

●一般的なことに関しての怒⇒宥める
 【ユ】本当に、頭にくるな。
 【ロ】まあまあ、抑えて抑えて。

【図10】



【図11】

:
 【ロ】 「Aさん、好きな食べ物は」
 【ユA】 「リンゴだよ」 ←(すぐに答えた)
 【ロ】 「Aさんはリンゴが大好きなんだね」
 【ロ】 「じゃあ、Bさんの好きな食べ物は」
 【ユB】 「えーとね、バナナかな」
 ←(10秒程度考えてから答えた)
 【ロ】 「Bさん、本当にバナナが好きなのかな」
 :
 :
 【ユA】 「ロボットさん、サッカーは何人でやるか知ってる」
 ←(知らないことを聞かれた)
 【ロ】 「分かんないな、教えてよ」
 ←(ユーザに問い返す)
 【ユA】 「サッカーはね、11人でやるんだよ、ロボットさん」
 【ロ】 「教えてくれてありがとう」
 ←(サッカーの人数の質問と答えを学習)
 :
 :
 【ユC】 「ところでロボット君、サッカーって何人でやるのかな」
 【ロ】 「サッカーは11人でやるよ」
 ←(前に一度聞かれていて学習したので答えることができた)
 【ユC】 「よく知ってるね」
 :

【書類名】 要約書

【要約】

【課題】 ユーザとの対話状況に応じて発話内容を臨機応変に変えることができ、知的で自然な音声対話を実現する音声対話装置を提供する。

【解決手段】 音声対話装置 1 においては、ユーザから知らないことを聞かれた場合にユーザにその答えを問い返し、その質問内容と答えを記憶して、次からの対話に用いるようにする。このため、知らない対話内容によって対話を中断させたり、ユーザの提示する話題を変更したりする必要性が小さくなると共に、学習によって新たなシナリオや語彙を増やして知識を向上させ、次回からのユーザとの対話に反映することができる。その結果、学習を重ねる毎に特定のユーザに対して満足のできる対話を実現することができるようになる。また、異なるユーザに対しては新たな話題や情報を提供することができ、知的な対話を実現することができる。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 4 2 6 0]

1. 変更年月日	1 9 9 6 年 1 0 月 8 日
[変更理由]	名称変更
住 所	愛知県刈谷市昭和町 1 丁目 1 番地
氏 名	株式会社デンソー